DOCUMENT RESUME

ED 442 917 UD 033 660

| | |
|---|---|
| AUTHOR | Johnson, Kirk A. |
| TITLE | Do Small Classes Influence Academic Achievement? What the National Assessment of Educational Progress Shows. A Report of the Heritage Center for Data Analysis. |
| INSTITUTION | Heritage Foundation, Washington, DC. |
| REPORT NO | CDA00-07 |
| PUB DATE | 2000-06-09 |
| NOTE | 16p. |
| AVAILABLE FROM | Heritage Foundation, 214 Massachusetts Avenue, N.E., Washington, DC 20002. Tel: 202-546-4400; Web site: http://www.heritage.org. |
| PUB TYPE | Reports - Research (143) |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | *Academic Achievement; *Class Size; Elementary Secondary Education; National Competency Tests; *Reading Achievement; Sex Differences; *Small Classes |
| IDENTIFIERS | National Assessment of Educational Progress |

ABSTRACT

This study used data from the 1998 National Assessment of Educational Progress (NAEP) reading examination to analyze the effect of class size on academic achievement. The NAEP is administered in 4th, 8th, and 12 grades to measure academic achievement in various fields. Besides providing data on academics, it offers information on student, teacher, and administrator background. This study analyzed six factors: class size; race and ethnicity; parents' educational attainment; number of reading materials in the home; free or reduced price lunch participation; and gender. Results indicate that on average, being in a small class does not increase the likelihood that a student will attain a higher score on the NAEP reading test. Children in the smallest classes do not score higher than students in the largest classes. An appendix presents results of the statistical models. (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 442 917

# A Report of The Heritage Center For Data Analysis

## Do Small Classes Influence Academic Achievement? What the National Assessment of Educational Progress Shows

Kirk A. Johnson, Ph.D.

CDA00-07 June 9, 2000

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

B. Plevn
Heritage Foundation

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

UD033660

The Heritage Foundation

# A Report of The Heritage Center for Data Analysis

## Do Small Classes Influence Academic Achievement? What the National Assessment of Educational Progress Shows

Kirk A. Johnson, Ph.D.

CDA00-07 June 9, 2000

The Heritage Foundation
214 Massachusetts Avenue, NE ´ Washington, DC 20002 ´ (202) 546-4400 ´ www.heritage.org

*NOTE: Nothing written here is to be construed as necessarily reflecting the views of The Heritage Foundation or as an attempt to aid or hinder the passage of any bill before Congress.*

# DO SMALL CLASSES INFLUENCE ACADEMIC ACHIEVEMENT? WHAT THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS SHOWS

KIRK A. JOHNSON, PH.D.

Do small classes make a difference in the academic achievement of elementary school students? From the attention given this subject by politicians, it would be reasonable to assume that class size has been shown to be essential to good academic outcomes. Congress, for example, allocated $1.3 billion for the "Class Size Reduction" provision of the Elementary and Secondary Education Act (ESEA) in fiscal year 2000. The Clinton Administration has requested even more funding for FY 2001.[1] And there are proposals to pump large sums of money into efforts to increase the number of teachers in public elementary schools in order to decrease the ratio of students to teachers.[2]

This report uses data from the 1998 National Assessment of Educational Progress (NAEP) reading examination to analyze the effect of class size on academic achievement. The NAEP provides the most comprehensive database on educational outcomes available to researchers. Among the major findings of this analysis of NAEP data are that:

- On average, being in a small class does not increase the likelihood that a student will attain a higher score on the NAEP reading test, and
- Children in the smallest classes (those with 20 or fewer students per teacher) do not score higher than students in the largest classes (those with 31 or more students per teacher).

## BACKGROUND

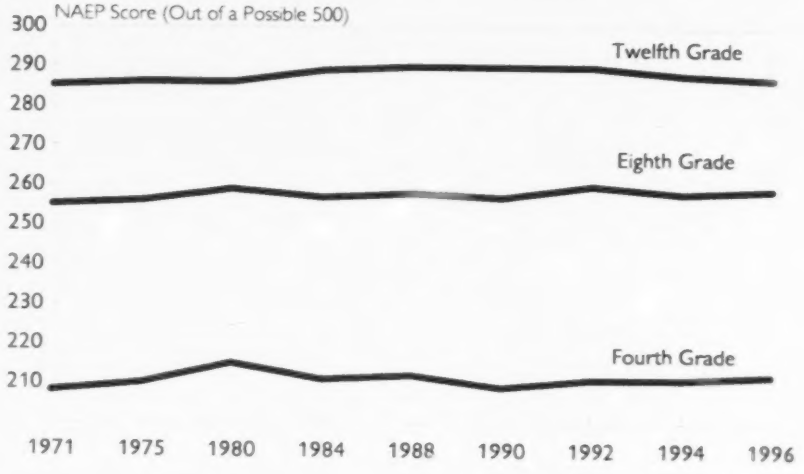
Most Americans believe that educating children in smaller classes would improve educational outcomes. Indeed, according to an NBC News/*Wall Street Journal* poll taken in March 1997, some 70 percent of adults believe that reducing class size would lead to significant academic improvements in public schools.[3]

1. U.S. Department of Education, "Total Appropriations for ESEA, 1990-2001," unpublished tables available upon request from the author.
2. The White House, "President Clinton Highlights Education Reform Agenda with Roundtable on What Works," press release, May 4, 2000, at *http://www.pub.whitehouse.gov/urires/I2R?urn:pdi://oma.eop.gov.us/2000/5/5/8.text.1*. See also The White House, "Remarks by the President in Roundtable on Reforming America's Schools," May 4, 2000, at *http://www.pub.whitehouse.gov/uri-res/I2R?urn:pdi://oma.eop.gov.us/2000/5/5/14.text.1*.
3. Hart and Teeter Research Companies, *NBC News/Wall Street Journal Poll*, March 1997; see Question 108, as cited in Eugene M. Lewit and Linda Schuurmann Baker, "Class Size," *The Future of Children*, Vol. 7 (1997), pp. 112–121.

But elementary and secondary school class sizes have fallen steadily over the past few decades. In 1970, public schools averaged 22.3 students per teacher nationwide. By the late 1990s, however, public schools averaged about 17 students per teacher, due to a combination of demographic trends and conscious policy decisions to lower the ratios.[4]

Chart 1 CDA00-07?

**National Assessment of Educational Progress Reading Scores: Average Scale, 1971–1996**




Source: National Assessment of Educational Progress Long Term Trend Reading Data, 1971–96

Over the same period, however, academic achievement, as measured by the NAEP exam, stayed relatively constant. Achievement for all three grades (fourth, eighth, and twelfth) that take the NAEP tests may vary slightly from year to year, but as shown in Chart 1, the average score on the reading test has changed very little over the past 25 years. At face value, this record of "stability" may not be sufficient evidence to conclude that the decline in class size has had no influence on test scores. It does, however, illustrate the trend in academic achievement over time in America's schools.[5]

The academic literature on the impact of low class size on academic achievement has been decidedly mixed. One of the most frequently cited reports on class size is Frederick Mosteller's study of young elementary school students in Tennessee.[6] Mosteller found a significant difference in achievement between the students in classes of 15 students per teacher and those in classes of 23. Recently, however, University of Rochester economist Eric Hanushek has questioned the results of this study, noting that "the bulk of evidence...points to no systematic effects of class size reductions within the relevant policy range."[7] Of the studies that do demonstrate some statistically significant gains in achievement, most generally involve substantial reductions in class size.[8] However, none of the current national policy proposals would massively shrink class size.[9] Clearly, more research is needed on this subject.

4. U.S. Bureau of the Census, *Statistical Abstract of the United States 1998* (Washington, D.C.: U.S. Government Printing Office, 1998), using data from National Center for Education Statistics, *Digest of Education Statistics* (1974 to 1998), published 1998.
5. University of Rochester economist Eric Hanushek does argue that if large class sizes are a problem today, they must have been a more serious problem in the past. See Eric Hanushek, "The Evidence on Class Size," in Susan E. Mayer and Paul Peterson, eds., *Earning and Learning: How Schools Matter* (Washington, D.C.: Brookings Institution, 1999), pp. 131–168.
6. Frederick Mosteller, "The Tennessee Study of Class Size in the Early School Grades," *The Future of Children*, Vol. 5 (1995), pp. 113–127.
7. Eric Hanushek, "Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects," *Educational Evaluation & Policy Analysis*, Vol. 21 (1999), p. 144.
8. R. Slavin, "Achievement Effects of Substantial Reductions in Class Size," in R. Slavin, ed., *School and Classroom Organization* (Hillside, N.J.: Erlbaum, 1989), pp. 247–257.

## HOW TO INTERPRET THESE FINDINGS

This report contains the results of statistical tests that use NAEP data to explain differences in reading test scores. These statistical tests isolate the independent effects of a number of factors on reading scores (such as the education of parents) in order to determine whether class size matters to these test scores. The statistical tests (or correlations) cover data on a wide array of school children, as defined by their race, income, and other socioeconomic characteristics. Because the statistical model used here includes these socioeconomic characteristics, the reader can interpret these findings as applicable to each of these groups of students. Thus, the findings about class size and reading scores apply as much to upper-income as to lower-income students, to blacks as to whites, to girls as to boys, and so forth.

These correlations suggest that there is a statistical relationship between the factor and achievement in reading, but they do not suggest that these independent factors *cause* differences in academic achievement.

The variables in the model came from the NAEP database and do not include everything that might have an effect on academic achievement, such as the methods used to teach reading. These factors may be much more important in general, or for a particular child, than the factors recorded in the NAEP data. Moreover:

- Some variables, such as participation in the federal free and reduced-price lunch program, are proxies (substitutes) for other unobserved factors. For example, eligibility for the free and reduced-price lunch program is determined by income; only children from low-income families may participate. Although not all low-income children will participate in the free and reduced-price lunch program, many will. Such information may be used, then, to analyze the effect of different characteristics on achievement.
- Some variables also may be used to determine the effect of some unobservable "third factor." For example, this model does not suggest that poor families have children who do worse on the NAEP *because they are poor*. Rather, poor families may have some unobservable characteristics or challenges that make it more difficult to succeed in school. Similarly, the categories of black and Hispanic students cover children whose characteristics other than their race may make it more difficult for them to score well.
- "Statistically insignificant" means that the effect of the variable/factor, if any, is no different from zero effect. For example, if the relationship between small class size and academic achievement is statistically insignificant, that means that students in small classes do no better than those in large ones.

## CHARACTERISTICS OF THE NAEP DATA

The author used the 1998 NAEP database of reading to measure the influence of class size on academic achievement. The NAEP, first administered in 1969, is an examination that measures academic achievement in a variety of fields, such as reading, writing, mathematics, science, geography, civics, and the arts. Currently, the NAEP is administered to fourth, eighth, and twelth grade students, with the main tests in math and reading

9. One of President Clinton's social policy objectives is the funding of 100,000 new teachers; these 100,000 new teachers will not significantly change the nationwide student–teacher ratio. According to the *Digest of Education Statistics*, there were some 46.8 million public school students and 2.8 million teachers in 1997, rendering a student–teacher ratio of 16.8 to 1. If these 100,000 teachers were hired tomorrow, it would only cause the national student–teacher ratio to drop by 3.5 percent. Mosteller's research (absent the Hanushek critique) suggests that class sizes would have to drop by one-third before significant gains in academic achievement would be found.



BEST COPY AVAILABLE

given alternately every two years. For example, reading was tested in 1998; math was assessed in 1996 and 2000.

The NAEP is actually two tests, a nationally administered test and state-administered tests. Over 40 states participate in the separate state samples used to gauge achievement within those individual jurisdictions. For the purposes of this study, only 1998 national reading data were used.

The most significant benefit of using the NAEP data is that, in addition to test scores in the subject area, it includes an assortment of background information for the students taking the exam, their main subject-area teacher, and their school administrator. Responses from the teachers and school administrators are linked to the student's information, which yields a rich database of information. The background questions include:

- TV viewing habits,
- Computer usage at home and school,
- Teacher tenure and certification,
- Socioeconomic status,
- Basic demographics, and
- School characteristics.

By incorporating this information with their assessment of the NAEP data, researchers can glean a great deal of evidence into the factors that explain the differences found in NAEP scores among children.

## THE HERITAGE ANALYSIS

This analysis looked at academic achievement by analyzing six factors: class size, race and ethnicity, parents' educational attainment, number of reading materials in the home, free or reduced-price lunch participation, and gender. Using regression analysis, Heritage analysts can isolate the effect of each factor. The Heritage analysis uses a jackknifed ordinary least squares model[10] and looks at the effects of these factors on the NAEP 1998 nationwide sample of public school children.[11]

### The Independent Variables

1. **Class Size.** Frederick Mosteller explains why small classes boost achievement: "Having fewer children in class reduces the distractions in the room and gives the teacher more time to devote to each child."[12] The average time a teacher can spend with each child, then, appears to be important in the learning process. To address class size, this analysis studies the NAEP data in two different ways (statistical models). The first compares the academic outcomes of children in the smallest classes (20 or fewer students per teacher) with those of all other students. The second only compares the children in these small classes with those in large classes (at least 31 students per teacher).

2. **Race and Ethnicity.** Many studies and reports have demonstrated that over time, African–American and Latino students tend to perform more poorly on standardized tests than do white students (although the gap has generally narrowed over the past 25 years).[13] There are a number of potential explanations for this trend.[14] Because strong differences exist in academic achievement among the races, the variables of race and ethnicity are included in the analysis.

10. Ordinary least squares is a general statistical regression technique that is often used by researchers. See Michael Lewis-Beck, *Applied Regression: An Introduction* (Beverly Hills, Cal.: Sage Publications, 1980), from Sage Publications' *Quantitative Applications in the Social Sciences*, Series No. 07–022. A jackknife is a complex resampling technique that is designed to accurately estimate statistical significance from surveys such as the NAEP that employ a complex sampling methodology. See Appendix A for the results and more information on the jackknifed ordinary least squares model.
11. Private school children are excluded from this analysis.
12. Mosteller, "The Tennessee Study of Class Size in the Early School Grades," p. 125.
13. For an analysis of the long-term achievement gap, see U.S. Department of Education, *Report in Brief: NAEP 1996 Trends in Academic Progress* (Washington, D.C.: U.S. Government Printing Office, 1997), Figure 2, p. 14.
14. One recent compilation on this subject is Christopher Jencks and Meredith Phillips, eds., *The Black–White Test Score Gap* (Washington, D.C.: Brookings Institution Press, 1998).

3. **Parents' Education.** Many researchers have noted that the educational attainment of a child's parents is a good predictor of their child's academic achievement. Parents who, for instance, are college educated could be better equipped to help their children with homework and understanding concepts than are those who have less than a high school education, all other things being equal. Because the education level of one parent is often highly correlated with the other's, only a single variable is included in the analysis.

Table 1 CDA00-07

**Average 1998 NAEP Reading Scores**
**Estimated Scores for One Hypothetical Base Case**

| | Model 1 All Students | Model 2 Small vs. Large Class Size Students |
|---|---|---|
| Fourth Grade | 223.6 | 217.6 |
| Eighth Grade | 260.6 | 26[illegible]6 |

Note: The maximum score on the NAEP test is 500. The base case is a white female not participating in the free/reduced price lunch program who has two reading materials in the home and is in a class larger than 20 students to one teacher.

4. **Number of Reading Materials in the Home.** The presence of books, magazines, encyclopedias, and newspapers generally indicates a dedication to learning in the household. Researchers have determined that these reading materials are important aspects of the home environment.[15] The analysis thus includes a variable controlling for the number of these four types of reading materials found at home.

5. **Free/Reduced-Price Lunch Participation.** Income is often a key predictor of academic achievement because low-income families seldom have the resources to purchase extra study materials or tutorial classes that may help their children perform better in school. While the NAEP does not collect data on household income, it does collect data on participation in the school free and reduced-price lunch program that are used here.[16]

6. **Gender.** Empirical research has suggested that girls tend to perform better on reading and writing subjects while boys perform better on the more analytical subjects of math and science.[17] Many authors have expounded on this idea,[18] yet the data on the male–female achievement gaps are often inconsistent. In 1998, for example, young men scored higher than young women on both the verbal and quantitative sections of the Scholastic Achievement Test (SAT). Some writers have noted that this may be because of a fundamental bias against females in the educational system.[19] Another explanation, however, is that the test results reflect a selection bias in which more "at-risk" females opt to take the SAT relative to males who take it.[20] In order to account for this factor, the analysis includes a variable for gender.

7. **Omitted variables.** Previous research[21] has included more family background variables in

15. Such opinions have been prevalent for years. See, for example, James S. Coleman, Thomas Hoffer, and Sally Kilgore, *High School Achievement* (New York: Basic Books, 1982).
16. Since eligibility for the free and reduced-price lunch program is determined by household income relative to the official poverty line, this variable is used as a good proxy for income.
17. U.S. Department of Education, *NAEP 1994 Trends in Academic Progress* (Washington, D.C.: U.S. Government Printing Office, 1996).
18. For a brief discussion of this point of view, see Thomas Hancock *et al.*, "Gender and Developmental Differences in the Academic Study Behaviors of Elementary School Children," *Journal of Experimental Education*, Vol. 65 (1996), pp. 18–39.
19. See Myra and David Sadker, *Failing at Fairness: How America's Schools Cheat Girls* (New York: Simon & Schuster, 1994).
20. The College Board, *1999 College Bound Seniors* (New York: The College Board, 1999).

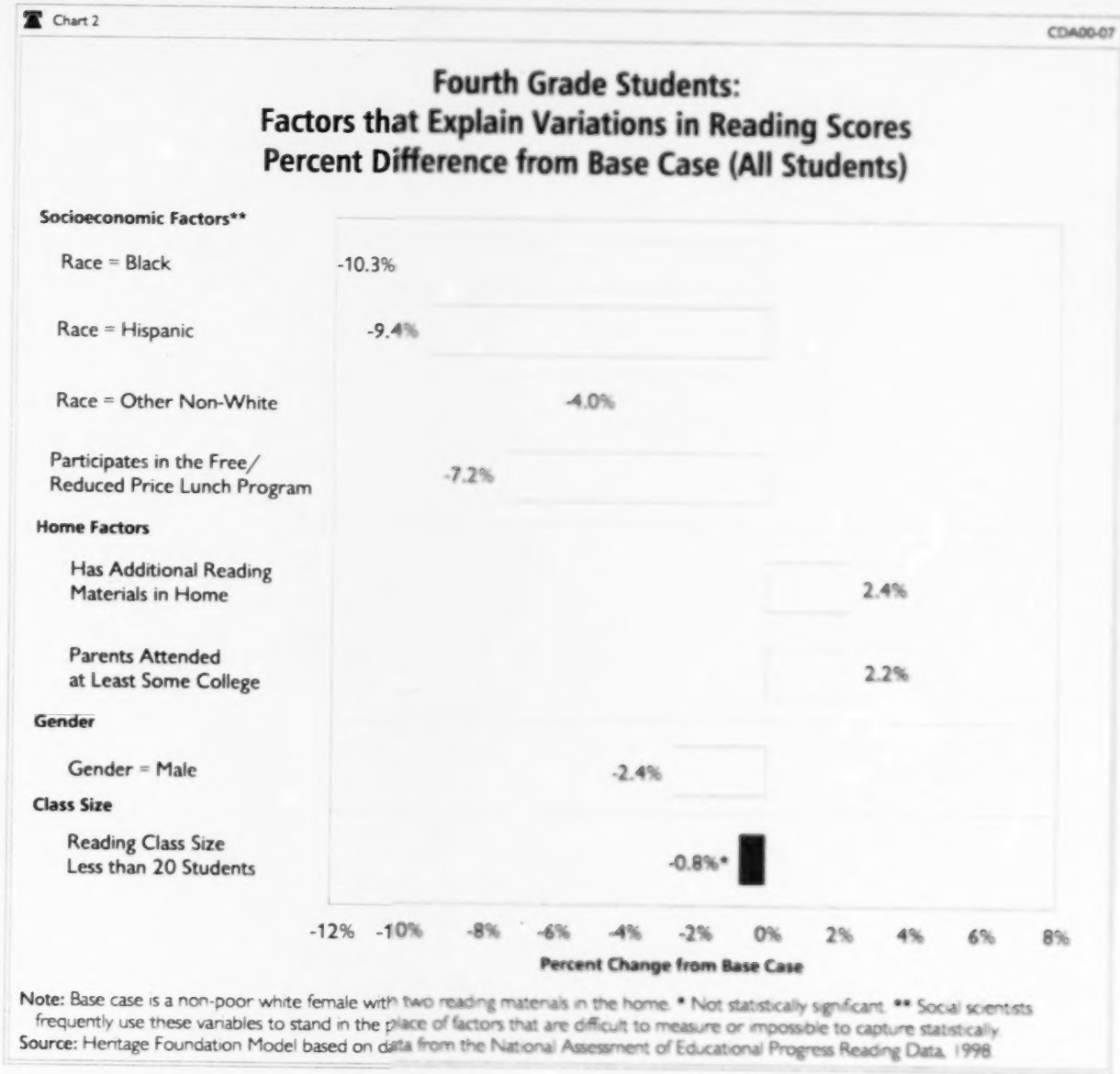


Chart 2
CDA00-07
Fourth Grade Students:
Factors that Explain Variations in Reading Scores
Percent Difference from Base Case (All Students)
Socioeconomic Factors**
Race = Black
-10.3%
Race = Hispanic
-9.4%
Race = Other Non-White
-4.0%
Participates in the Free/ Reduced Price Lunch Program
-7.2%
Home Factors
Has Additional Reading Materials in Home
2.4%
Parents Attended at Least Some College
2.2%
Gender
Gender = Male
-2.4%
Class Size
Reading Class Size Less than 20 Students
-0.8%*
-12% -10% -8% -6% -4% -2% 0% 2% 4% 6% 8%
Percent Change from Base Case
Note: Base case is a non-poor white female with two reading materials in the home. * Not statistically significant. ** Social scientists frequently use these variables to stand in the place of factors that are difficult to measure or impossible to capture statistically.
Source: Heritage Foundation Model based on data from the National Assessment of Educational Progress Reading Data, 1998.

the model specification. In the 1998 NAEP database, however, the only information available on children's parents is their educational attainment. The NAEP does not ask whether the child lives with both parents (or parental figures), one parent, or no parents (i.e., in a group home). Future administrations of the NAEP test should include this type of question since a great deal of research has found that having both parents in the home can improve a child's academic achievement.

## RESULTS OF THE ANALYSIS

These six factors formed the basis of two statistical models[22] that were applied to the NAEP's 1998 nationwide sample of public school children who took the reading test.[23] As noted above, the first

21. See, for example, Kirk A. Johnson, "Comparing Math Scores of Black Students in D.C.'s Public and Catholic Schools," Heritage Foundation *Center for Data Analysis Report* No. CDA99-08, October 7, 1999.
22. See Appendix A for the results and a more complete discussion of the jackknifed ordinary least squares model.
23. Private school children were excluded from this analysis.

Chart 3 CDA00-07

## Eighth Grade Students: Factors that Explain Variations in Reading Scores Percent Difference from Base Case (All Students)

Note: Base case is a non-poor white female with two reading materials in the home. * Not statistically significant. ** Social scientists frequently use these variables to stand in the place of factors that are difficult to measure or impossible to capture statistically.
Source: Heritage Foundation Model based on data from the National Assessment of Educational Progress Reading Data, 1998.

model compares the data for children in small class sizes (20 or less students per teacher) to all other students. The second model only compares data for students reported to be in either small or large classes (classes with 31 or more students per teacher). By determining whether or not an achievement difference exists between the smallest and largest classes in America, the second model addresses the contention that there may be differences in achievement as the class size gap widens.

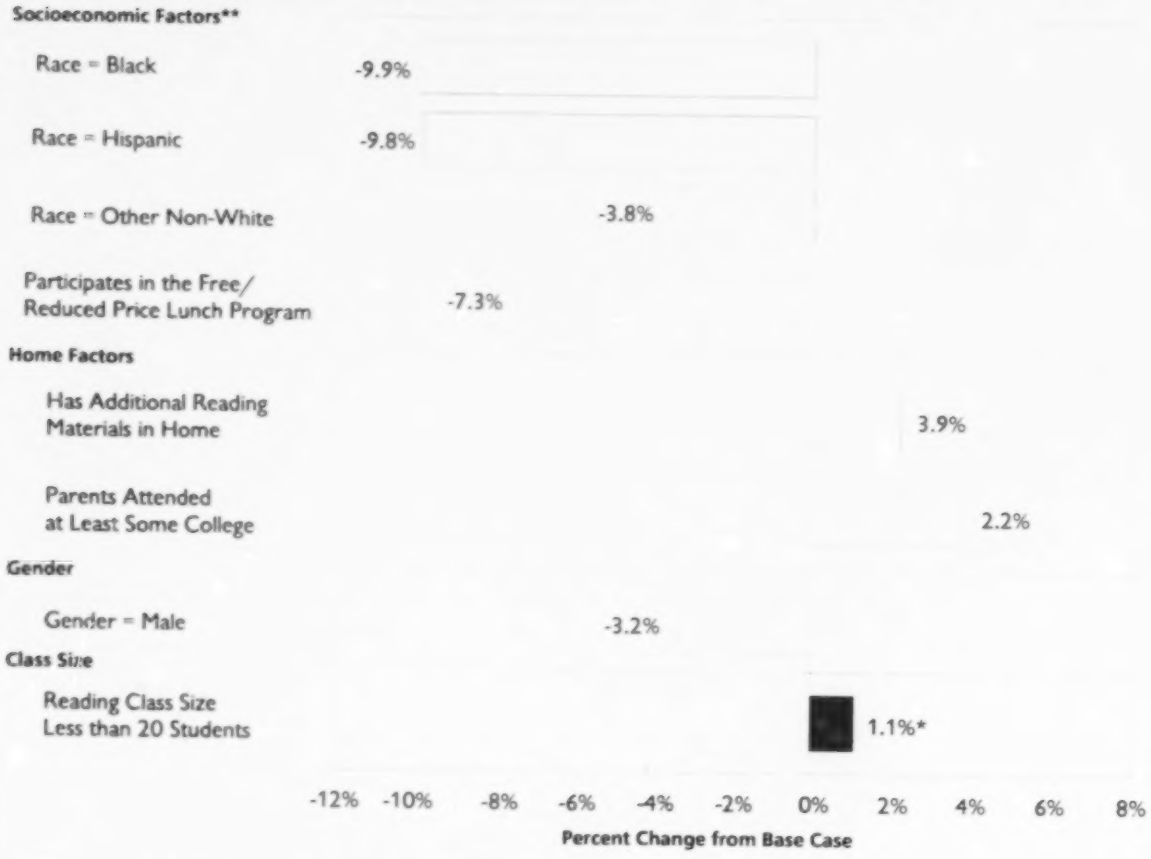
Chart 2 and Chart 3 show the percent change in fourth and eighth grade reading scores attributable to the factors in the first model, compared with a base case, while Chart 4 and Chart 5 show the percent change in the second model.[24] Here, the base case is defined as a child with the following characteristics:

- White;
- Female;
- Non-poor (that is, not participating in the free

24. Specifying a base case with which to assess the results of a regression model is fairly arbitrary. Changing the base model case does not alter the interpretation of the results.

Chart 4 CDA00-07

## Fourth Grade Students: Factors that Explain Variations in Reading Scores Percent Difference from Base Case (Small versus Large Classes)




Note: Base case is a non-poor white female with two reading materials in the home. * Not statistically significant. ** Social scientists frequently use these variables to stand in the place of factors that are difficult to measure or impossible to capture statistically.
Source: Heritage Foundation Model based on data from the National Assessment of Educational Progress Reading Data, 1998.

and reduced price lunch program);
- Parents who did not attend college;
- Has two out of the four possible reading materials in the home; and
- Has a reading class size of over 20 students to one teacher.

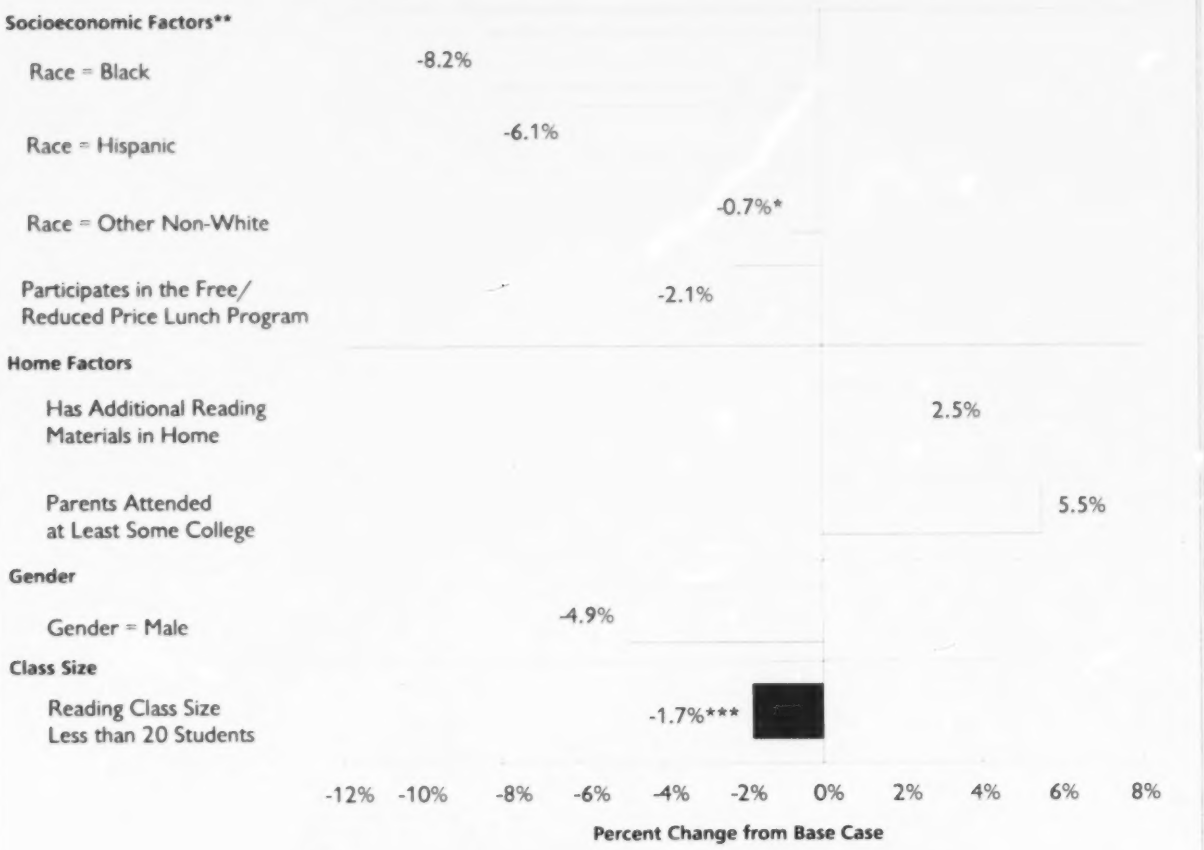
The estimates of the base case are reported in Table 1 for both models. These are the scores that a hypothetical individual would score out of a maximum possible NAEP score of 500. Chart 2 through Chart 5 show the positive or negative percent changes for each variable, holding constant all other variables in the model.

In the first model, the analysis of the data on children in all class sizes shows no significant difference in reading test scores attributable to class size, holding all other variables constant.[25] As seen in Chart 2, NAEP scores of fourth grade children

25. The variables of race and ethnicity, parental college attendance, poverty, gender, and reading materials in the home are held constant.

Chart 5 CDA00-07

## Eighth Grade Students: Factors that Explain Variations in Reading Scores Percent Difference from Base Case (Small versus Large Classes)




**Note:** Base case is a non-poor white female with two reading materials in the home. * Not statistically significant. ** Social scientists frequently use these variables to stand in the place of factors that are difficult to measure or impossible to capture statistically. *** Barely statistically significant.
**Source:** Heritage Foundation Model based on data from the National Assessment of Educational Progress Reading Data, 1998.

whose parents attended some college are 2.2 percent higher than scores for children whose parents have a high school education or less. Most important, moving from a class size above 20 down to 20 or fewer reduces NAEP scores by 0.8 percent, but this effect is statistically indistinguishable from no influence. While it may seem logical that lower class sizes would have a positive influence on achievement, the NAEP data do not support that conclusion. The second model, comparing children in small classes to those in large classes, reaches a similar conclusion. Again, class size does not have a meaningful impact on academic achievement.

For fourth graders, the model results do not change appreciably when comparing only those in small or large classes. One exception is the variable that controls for having at least one parent who attended college. The importance of this variable increases when the model compares children in small and large class sizes.

For eighth graders, the class size variable is significant when comparing children in small and large classes. The results of the comparison are counterintuitive since the coefficient has a negative sign. Holding other variables constant, this means that eighth grade children in small class sizes do worse on the NAEP reading exam than do those in large classes. The magnitude of the effect is significant; in the base model, a child would score 1.7 percent less than the base case child. The variable is barely significant statistically,[26] however, and should be treated with suspicion.

Both fourth and eighth grade girls score slightly higher than do boys on the reading exam, which bolsters recent evidence on gender differences in academic achievement. Girls on average, notes American Enterprise Institute W. H. Brady Fellow Christina Hoff Sommers, "get better grades, are more engaged academically, and are now in the majority in higher education."[27] The results here support the contention that schools are not short-changing girls.[28]

## CONCLUSION

Class size has little or no effect on academic achievement, according to this analysis of 1998 NAEP data. It is quite likely, in fact, that class size as a variable pales in comparison with the effects of many factors not included in the NAEP data, such as teacher quality and teaching methods. Observes Irwin Kurz, principal of the highly successful P.S. 161, a public school in Brooklyn, New York, that serves poor children and has an average class size of 35, it is "[b]etter to have one good teacher, than two crummy teachers any day."[29]

*—Kirk A. Johnson, Ph.D., is a Policy Analyst in the Center for Data Analysis at The Heritage Foundation.*

26. In technical terms, the t-test on the class size coefficient has a significance level that comes close to .05, or 5 percent. In light of the other results, and since one would expect those in smaller classes to perform better, researchers might question this result; however, that judgment is left to the reader.
27. Christina Hoff Sommers, "The War Against Boys," *The Atlantic Monthly*, Vol. 285 (May 2000), p. 60.
28. Recent publications continue to advance the argument that schools, through accident or design, limit the success of girls. See, for example, American Association of University Women, ed., *Gender Gaps: Where Schools Still Fail Our Children* (New York: Marlowe & Co., 1998).
29. See Samuel Casey Carter, *No Excuses: Lessons from 21 High-Performing, High-Poverty Schools* (Washington, D.C.: The Heritage Foundation, 2000), pp. 74–77.

## APPENDIX A: RESULTS OF THE STATISTICAL MODELS

Table 2 and Table 3 report the results of an analysis of NAEP data using two statistical models. Table 2 shows the coefficients and significance tests for the first model, which compares data for all public school children in the analysis, while Table 3 reports the results for the model that compares only small classes (20 students or less per teacher) to large classes (31 students or more per teacher). As shown in these tables, most variables are statistically significant.[30] Contrary to conventional wisdom, the class size variable is not significant or has the wrong sign on the coefficient.

Table 2 CDA00-07

### Regression Analysis of NAEP Reading Achievement for Public School Students Nationwide All Children Entered in Model

**Fourth Grade Reading Score Model: Effects of Independent Variables on Reading Score**

| | Coefficient | T-Test | Significance |
|---|---|---|---|
| (Constant)[1] | 212.822 | 86.319 | 0.0000 |
| Race = Black | -23.117 | -11.138 | 0.0000 |
| Race = Hispanic | -21.062 | -9.343 | 0.0000 |
| Race = Other Non-White | -9.039 | -3.182 | 0.0015 |
| Parents Attended at Least Some College | 4.892 | 3.097 | 0.0020 |
| Has Additional Reading Materials in Home | 5.383 | 7.834 | 0.0000 |
| Participates in the Free/Reduced Price Lunch Program | -16.045 | -10.668 | 0.0000 |
| Gender = Male | -5.424 | -4.863 | 0.0000 |
| Class Size = 20 Students to 1 Teacher or Less | -1.845 | -0.857 | 0.3918 |
| Explanatory Power: $R^2$ = 0.2352 | | | |

**Eighth Grade Reading Score Model: Effects of Independent Variables on Reading Score**

| | Coefficient | T-Test | Significance |
|---|---|---|---|
| (Constant)[1] | 247.439 | 107.339 | 0.0000 |
| Race = Black | -22.123 | -11.843 | 0.0000 |
| Race = Hispanic | -13.298 | -6.455 | 0.0000 |
| Race = Other Non-White | -1.381 | -0.505 | 0.6135 |
| Parents Attended at Least Some College | 13.501 | 9.748 | 0.0000 |
| Has Additional Reading Materials in Home | 6.570 | 10.577 | 0.0000 |
| Participates in the Free/Reduced Price Lunch Program | -6.406 | -5.232 | 0.0000 |
| Gender = Male | -12.478 | -11.036 | 0.0000 |
| Class Size = 20 Students to 1 Teacher or Less | -2.158 | -1.169 | 0.2427 |
| Explanatory Power: $R^2$ = 0.2787 | | | |

Note: 1. The "constant" term refers to the reading score value when all other model variables.
Source: Data are from the 1998 National Assessment of Educational Progress Reading Exam.

In this analysis, there are two statistical issues to consider. First, the NAEP exam is a long test and therefore is not administered in its entirety to all children. Rather, different parts are given to different children. Certain students will do better on certain portions of the test than others. Consequently, a "true" score must be estimated, or imputed, from the incomplete information. The NAEP estimates five plausible composite reading scores and recommends that researchers use all five in any analysis. The Heritage analysis described here follows the guidelines specified by the Educational Testing Service (which works closely with the National Center for Education Statistics in developing the data file) for incorporating all five reading scores into the analysis.[31]

Second, the NAEP utilizes a complex sample design, oversampling children with certain characteristics.[32] Each child, then, is given a unique weight, which is calculated from the probability of being selected from the population at large (in this

30. This means that these variables have no statistically discernable difference between the coefficient value and zero, so there is no effect. Statistical significance is usually pegged at a 5 percent or 10 percent level. See Lewis-Beck, *Applied Regression: An Introduction*.

Table 3 CDA00-07

## Large Classes Compared to Small Classes: Regression Analysis of NAEP Reading Achievement for Public School Students Nationwide

**Fourth Grade Reading Score Model: Effects of Independent Variables on Reading Score**

| | Coefficient | T-Test | Significance |
|---|---|---|---|
| (Constant)[1] | 208.124 | 34.950 | 0.0000 |
| Race = Black | -21.452 | -6.062 | 0.0000 |
| Race = Hispanic | -21.286 | -5.117 | 0.0000 |
| Race = Other Non-White | -8.184 | -1.659 | 0.0973 |
| Parents Attended at Least Some College | 8.420 | 2.502 | 0.0124 |
| Has Additional Reading Materials in Home | 4.761 | 4.230 | 0.0000 |
| Participates in the Free/Reduced Price Lunch Program | -15.961 | -5.219 | 0.0000 |
| Gender = Male | -6.912 | -3.392 | 0.0007 |
| Class Size = 20 Students to 1 Teacher or Less | 2.367 | 0.707 | 0.4797 |
| Explanatory Power: $R^2 = 0.2313$ | | | |

**Eighth Grade Reading Score Model: Effects of Independent Variables on Reading Score**

| | Coefficient | T-Test | Significance |
|---|---|---|---|
| (Constant)[1] | 249.211 | 61.022 | 0.0000 |
| Race = Black | -21.436 | -10.917 | 0.0000 |
| Race = Hispanic | -16.045 | -4.772 | 0.0000 |
| Race = Other Non-White | -1.945 | -0.526 | 0.5987 |
| Parents Attended at Least Some College | 14.326 | 5.814 | 0.0000 |
| Has Additional Reading Materials in Home | 6.692 | 5.622 | 0.0000 |
| Participates in the Free/Reduced Price Lunch Program | -5.465 | -2.630 | 0.0086 |
| Gender = Male | -12.854 | -6.938 | 0.0000 |
| Class Size = 20 Students to 1 Teacher or Less | -4.584 | -1.985 | 0.0472 |
| Explanatory Power: $R^2 = 0.2884$ | | | |

Note: 1. The "constant" term refers to the reading score value when all other model variables.
Source: Data are from the 1998 National Assessment of Educational Progress Reading Exam.

case, from the U.S. population of fourth or eighth graders in public schools). The NAEP's sample design requires a complex modeling technique, which the Heritage model employs.[33]

31. From a multivariate regression perspective, the model below must be replicated five times using each of the plausible values individually and then averaging the resulting coefficients to yield the final model results. In technical terms, this process corrects for measurement error in the reading score variable, since the test administrators do not actually observe the test score from taking the exam in its entirety.
32. For example, the NAEP typically oversamples for race and geography of school attended (e.g., urban, rural).
33. A procedure called a jackknife must be employed to correctly assess the variance of each variable's coefficient, and the NAEP database has a series of 62 "replicate weights" to aid in this task. These 62 jackknifes must be applied and the variances of each coefficient averaged for each of the five plausible test score models above (yielding a total of 315 models compiled for the purpose of this research). The WesVar Complex Samples software (produced by SPSS, Inc.) did much of this replication work. Using the jackknife results with the five plausible test score models allows for a variance correction mechanism. The purpose of the jackknife is to estimate a true sampling error. Correcting for the two types of error (measurement and sampling) allows for the most accurate estimates possible. See Bradley Efron, *The Jackknife, the Bootstrap, and Other Resampling Plans* (Philadelphia: Society for Industrial and Applied Mathematics, 1982), and Jun Shao and Dongsheng Tu, *The Jackknife and Bootstrap* (New York: Springer Verlag, 1995), for a more complete discussion of how this jackknife technique works.

# THE HERITAGE CENTER FOR DATA ANALYSIS

The Heritage Foundation's Center for Data Analysis provides timely and accurate analytical products intended to help advance public policy debates. To carry out this work, the Center employs the highest quality databases and maintains a set of customized and peer-reviewed analytical models.

The Center has created one of the largest privately held public policy databases in the United States: the Heritage Matched Database. This database statistically matches the federal government's principal public databases, and it allows Heritage analysts to study the effects of public policy changes on thousands of different types of families and individuals.

Center economists use specially developed models of federal tax policy, Social Security, and a wide range of other major programs to estimate how large and small policy changes will affect the federal budget and the pocketbooks of ordinary Americans. The Center shares its analysis of proposed legislation and administrative actions with policymakers in and out of government, providing members of Congress, officials within the executive branch, and the public policy community with an independent assessment of policy options.

The Center estimates the economic effects of policy changes through an integrated set of econometric models developed by WEFA, Inc., and maintained by the Center. Heritage analysts use the award-winning WEFA U.S. Macroeconomic Model to gauge a policy's effects on the national economy and on households. The Center traces these national effects to impacts on specific types of businesses through the WEFA U.S. Industry Model, and estimates how a policy will affect income, employment, and other aspects of economic life in each state through the 51 separate WEFA state models.

214 Massachusetts Ave., N.E. • Washington, D.C. 20002 • (202) 546-4400 • [illegible]